PS ManikantaVijayaSri NarasimhaNaidu Chikkala

+1 (240)-921-8135 | narasimhachikkala99@gmail.com | LinkedIn: narasimha-naidu-chikkala

EDUCATION

University of Maryland Baltimore County
Master's in Data Science 2022 - 2024

Jawaharlal Nehru Technological University
Bachelor's in Computer Science Engineering 2018 - 2022

TECHNICAL SKILLS

Web Technologies:	HTML, CSS, JavaScript, Bootstrap
Languages:	Java, Python
Version Control:	Git
ETL Tools:	Apache Airflow, AWS Glue, Informatica
Cloud Platforms:	AWS (Redshift, S3, EMR, Lambda)
Data Visualization Tools:	Tableau, Power BI
Database:	MySQL, PostgreSQL, SQL Server, Oracle, MongoDB
Data Modeling:	Dimensional modeling, Star/Snowflake Schema, ER Diagrams
Big Data Frameworks:	Hadoop, Apache Spark, MapReduce, Hive
Machine Learning:	Supervised learning (regression, classification), unsupervised learning (clustering, dimensionality reduction), Neural Networks, Data Mining
Libraries:	Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, TensorFlow, Keras, Statsmodels, NLTK, spaCy, PyTorch

PROFESSIONAL EXPERIENCE

SForce Corp. Texas
Data Engineer July 2023 – Present

- Managed large-scale data storage and retrieval from AWS S3, optimizing storage and ensuring efficient data accessibility for large datasets across distributed systems.
- Designed and implemented ETL pipelines using Apache Airflow, AWS Glue, and Informatica, to automate data extraction, transformation, and loading from multiple sources, including SQL databases, external APIs, and unstructured data.
- Developed and optimized data management processes, reducing data processing time for reporting and analytics by 30%.
- Launched and configured over 50 AWS EC2 instances, ensuring tailored security and networking settings, while administering scalable EMR clusters for processing data volumes of up to hundreds of gigabytes.
- Maintained service scalability, adapting to fluctuating demands (30-40% monthly) using AWS Glue to orchestrate ETL workflows for data volumes between 100GB to 1TB.
- Automated complex workflows using Apache Airflow, integrating AWS Lambda and Redshift, improving operational efficiency by 30% and streamlining end-to-end data processing.
- Developed and maintained real-time data processing pipelines using Kafka and Spark Streaming, enabling immediate availability of insights for critical business operations.
- Collaborated with cross-functional teams to develop dashboards and reports using Tableau and Power BI, ensuring real-time insights for decision-making and data security through AWS IAM for access control.
- Utilized version control systems (e.g., Git, GitHub) and implemented CI/CD pipelines to efficiently deploy and maintain data infrastructure changes with high reliability.
- Supported database maintenance and optimization across multiple systems, performing data validation checks and resolving discrepancies to ensure data quality and performance.
- Implemented data modeling techniques, such as Dimensional Modeling, Star/Snowflake Schema, and ER Diagrams, to design robust data architectures for improved querying and reporting.
- Leveraged Big Data frameworks like Hadoop, Apache Spark, MapReduce, and Hive to manage and process large-scale datasets, driving key business insights.

TechCiti Technologies Private Limited. India
Data Engineer Intern Jan 2022 – Jun 2022

- Collaborated with data scientists to provide clean, structured datasets for machine learning model training, leveraging ETL workflows to automate data preparation and integration.
- Utilized cloud-based monitoring tools such as AWS CloudWatch to track system performance, proactively identifying issues and minimizing downtime in the data infrastructure.
- Contributed to capacity planning and resource optimization, reducing cloud infrastructure costs by optimizing S3 storage classes and EC2 instance usage.
- Assisted in the development of automated testing suites for data pipelines, ensuring data quality and accuracy through validation and consistency checks before deployment.
- Optimized query performance for large-scale datasets by tuning SQL queries and indexing strategies, improving response times for critical analytics reports.
- Contributed to schema evolution for databases and data lakes, ensuring backward compatibility while adapting to evolving data models and business requirements.
- Tested and validated data pipelines to ensure they met performance benchmarks and followed best practices for error handling and minimizing failure points.

Verzeo Edutech Pvt. Ltd India
Data Science Intern Apr 2021 – Jun 2021

- Assisted in gathering, cleaning, and preprocessing large datasets using Pandas and NumPy for various data analysis tasks and machine learning models, ensuring high data quality and readiness for analysis.
- Developed and implemented predictive models using Python and libraries such as Scikit-learn, TensorFlow, and Keras to identify trends and support data-driven decision-making, including both supervised (regression, classification) and unsupervised (clustering, dimensionality reduction) learning techniques.
- Conducted exploratory data analysis (EDA) utilizing Matplotlib and Seaborn to uncover patterns and relationships in the data, providing actionable insights and visualizations to the team.
- Collaborated with cross-functional teams to deliver impactful data visualizations and dashboards using Tableau and Power BI, enhancing the accessibility of data insights for stakeholders.
- Presented findings and insights to key stakeholders, enabling better strategic decisions based on data, and leveraging insights drawn from data mining techniques.
- Contributed to the continuous enhancement of classification models through iterative refinement and hyperparameter tuning, culminating in a notable 15% increase in classification accuracy by utilizing advanced techniques such as neural networks and ensemble methods.

ACADEMIC PROJECTS

Flight Delay Analysis and Prediction | **Apache PySpark, MongoDB, Tableau**
- Utilized Apache Spark with MongoDB to analyze and predict flight delays based on historical data and weather patterns.
- Identified airlines with the most delays and potential causes and developed a Tableau dashboard for visual delay analysis.

Road Accident Severity Prediction Project | **Python, Flask, AWS**
- The goal of this project is to create an application that can assess and predict the level of severity for any road accidents.
- This project involves training a predictive model by balancing the historical accident data using SMOTE technique.
- Built and deployed a Flask-based web application on AWS EC2 for real-time road accident severity assessment.

Stock vs Crypto Price Prediction | **Python, Time Series Forecasting, Long Short-term Memory (LSTM)**
- Executed a comparative analysis predicting BTC-USD and S&P 500 stock prices using ARIMA and LSTM models.
- Applied statistical tests, differencing techniques, and autocorrelation analysis to ensure accurate modeling by stationarizing time series data.
- Demonstrated LSTM's superior forecasting performance over ARIMA, providing actionable insights for informed investment decisions that led to a 90% improvement in portfolio performance.

Research Paper

Blind Assistance System with Object Detection and Navigation May 2022

- Developed an object detection system with voice assistance using OpenCV and a pre-trained SSD model on COCO datasets for real-time object recognition.
- Implemented audio feedback via Pyttsx3 for object identification and obstacle warnings, creating a portable solution for visually impaired users.
- Submitted a research paper in GIS Science Journal under the ISSN No: 1869-9391.

CERTIFICATIONS

- AWS Certified Cloud Practitioner
- Java Fundamentals – Oracle